

ISDS 2014 Conference Abstracts



Digital Disease Detection Dashboard: Rapid Detection & Outbreak Management Tool

Karina N. Alvarez*, Catherine Ordun, Jane Blake, Kirsten A. Simmons, Keith Hansen, Dan Baker, Lynda Rowe, Yusra Ahmad, Donald M. Eby, Dimitrios Koutsonanos, Steve Escaravage and KC Decker

Booz Allen Hamilton, Atlanta, GA, USA

Objective

To develop a web-enabled Digital Disease Detection Dashboard (D4) that allows users to statistically model and forecast multiple data streams for public health biosurveillance. D4 is a user-friendly, cloud-enabled, and R Shiny-powered application that provides intuitive visualization enabling immediate situational awareness through interactive data displays and multi-factor analysis of traditional and non-traditional data feeds. The objective of D4 is to support public health decision making with high confidence across all four aspects of the biosurveillance continuum—detection, investigation, response, and prevention.

Introduction

Booz Allen Hamilton is developing a novel bio-surveillance prototype tool, the Digital Disease Detection Dashboard (D4) to address the questions fundamental to daily biosurveillance analysis and decision making: is something unusual happening (e.g., is an outbreak or novel disease emerging)?, What is the probability that what I'm seeing is by chance?, How confident am I that this data is really detecting a signal?, Why is this happening and can I explain it?; and How many cases should I expect? (e.g., magnitude of event over time). These questions focus on detection, confidence, variance, and forecasting and D4 integrates a number of diverse analytical tools and methods that are crucial to a complete biosurveillance program.

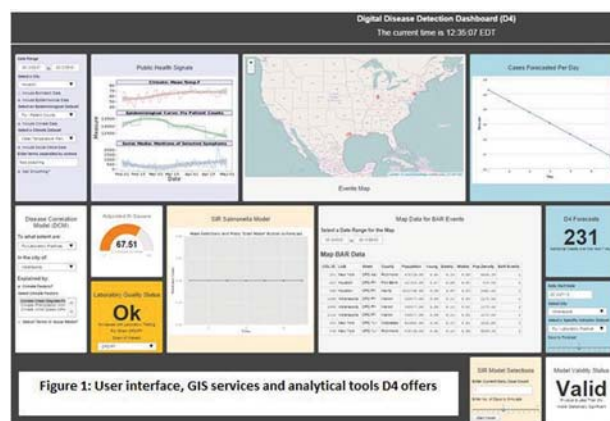
Description

D4 utilizes a big data cloud-based reference architecture and is comprised of widgets - each built in R Shiny Studio using R code. D4 includes a user interactive map with "hot spots" of potential adverse events, customized databases with rapid query data visualizations, searchable social media (Twitter) using regular stacked expressions, and forecast models. Central to D4 is the Disease Correlation Model (DCM), which uses linear regression to test the statistical significance of the relationship between the outcome of interest and explanatory factors (e.g., city, climate, and social media data). Included is an adjusted R^2 widget and p-value output to quickly quantify a DCM output goodness-of-fit. In addition, a forecasted widget calculates the number of estimated cases per day, over the number of days in the future by using estimates from a linear regression model that only includes time as the explanatory factor. For disease detection data generated from high-throughput laboratory analyses (e.g., real time PCR detection assays), a Laboratory Quality Status widget evaluates all associated data generated through quality management systems and outputs a Boolean (TRUE/FALSE) test result to provide the operator with a rapid quality assurance validation to laboratory data. D4 integrates multiple data streams into an R coding environment that enables open-source and plug-play statistical and widget-development capabilities. D4 backend resides in an Amazon cloud instance allowing for scalability, speed, and controlled distributed access. As a prototype, D4 currently injects examples of the multiple data streams that may be of interest such as social media (Twitter), climate data (precipitation, temperature), 911 call data, and two

explanatory data sets that are epidemiological data (foodborne illness cases, flu patient counts) and other laboratory diagnosis data. D4 is currently modeling three US cities and its architecture enables access and filtering for hundreds of historic and real-time data feeds via representational state transfer (REST) application programming interface (API) calls, enabling incorporation of standardized, reliable, and timely information.

Audience Engagement

D4 is bringing new innovations and analytical tools in the area of bio-surveillance, disease detection, and disease reporting. Depending on the source and quality of data, it can provide forecasting models with high accuracy accounting for different parameters and characteristics different types of disease demonstrate. Besides the advanced analytical tools D4 offers, what makes this platform distinct and innovative is the ability to accept and handle data from multiple sources and formats and allows for increased sensitivity and specificity of the system.



Keywords

Biosurveillance; Disease detection; Open-Source Analytics; Visualizations; R

Acknowledgments

JR Charlton and Kris Ford

*Karina N. Alvarez

E-mail: alvarez_karina@bah.com

